LA-UR-15-23327 (Accepted Manuscript)

On the universal structure of human lexical semantics

Youn, Hyejin
Bhattacharya, Tanmoy
Sutton, Logan
Smith, Eric
Moore, Christopher
Wilkins, Jon F.
Maddieson, Ian
Croft, William

Provided by the author(s) and the Los Alamos National Laboratory (2016-09-28).

To be published in: Proceedings of the National Academy of Sciences

DOI to publisher's version: 10.1073/pnas.1520752113

Permalink to record: http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-15-23327

Disclaimer:
Approved for public release. Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

On the universal structure of human lexical semantics

Hyejin Youn[a,b,c,1], Logan Sutton[d], Eric Smith[c,e], Cristopher Moore[c], Jon F. Wilkins[c,f], Ian Maddieson[g,h], William Croft[g], and Tanmoy Bhattacharya[c,i,1]

[a]Institute for New Economic Thinking at the Oxford Martin School, Oxford OX2 6ED, United Kingdom; [b]Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom; [c]Santa Fe Institute, Santa Fe, NM 87501; [d]American Studies Research Institute, Indiana University, Bloomington, IN 47405; [e]Earth-Life Sciences Institute, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550, Japan; [f]Ronin Institute, Montclair, NJ 07043; [g]Department of Linguistics, University of New Mexico, Albuquerque, NM 87131; [h]Department of Linguistics, University of California, Berkeley, CA 94720; and [i]MS B285, Grp T-2, Los Alamos National Laboratory, Los Alamos, NM 87545

Edited by E. Anne Cutler, University of Western Sydney, Penrith South, New South Wales, and approved December 14, 2015 (received for review October 23, 2015)

How universal is human conceptual structure? The way concepts are organized in the human brain may reflect distinct features of cultural, historical, and environmental background in addition to properties universal to human cognition. Semantics, or meaning expressed through language, provides indirect access to the underlying conceptual structure, but meaning is notoriously difficult to measure, let alone parameterize. Here, we provide an empirical measure of semantic proximity between concepts using cross-linguistic dictionaries to translate words to and from languages carefully selected to be representative of worldwide diversity. These translations reveal cases where a particular language uses a single "polysemous" word to express multiple concepts that another language represents using distinct words. We use the frequency of such polysemies linking two concepts as a measure of their semantic proximity and represent the pattern of these linkages by a weighted network. This network is highly structured: Certain concepts are far more prone to polysemy than others, and naturally interpretable clusters of closely related concepts emerge. Statistical analysis of the polysemies observed in a subset of the basic vocabulary shows that these structural properties are consistent across different language groups, and largely independent of geography, environment, and the presence or absence of a literary tradition. The methods developed here can be applied to any semantic domain to reveal the extent to which its conceptual structure is, similarly, a universal attribute of human cognition and language use.

polysemy | human cognition | semantic universals | conceptual structure | network comparison

The space of concepts expressible in any language is vast. There has been much debate about whether semantic similarity of concepts (i.e., the layout of this space) is shared across languages (1–9). On the one hand, all human beings belong to a single species characterized by, among other things, a shared set of cognitive abilities. On the other hand, the 6,000 or so extant human languages spoken by different societies in different environments across the globe are extremely diverse (10–12). This diversity reflects accidents of history as well as adaptations to local environments. Notwithstanding the vast and multifarious forms of culture and language, most psychological experiments about semantic universality have been conducted on members of Western, educated, industrial, rich, democratic (WEIRD) societies, and it has been questioned whether the results of such research are valid across all types of societies (13). The fundamental problem of quantifying the degree to which conceptual structures expressed in language are due to universal properties of human cognition, as opposed to the particulars of cultural history or the environment inhabited by a society, remains unresolved.

A resolution of this problem has been hampered by a major methodological difficulty. Linguistic meaning is an abstract construct that needs to be inferred indirectly from observations, and hence is extremely difficult to measure. This difficulty is even more apparent in the field of lexical semantics, which deals with how concepts are expressed by individual words. In this regard, meaning contrasts both with phonetics, in which instrumental measurement of physical properties of articulation and acoustics is relatively straightforward, and with grammatical structure, for which there is general agreement on a number of basic units of analysis (14). Much lexical semantic analysis relies on linguists' introspection, and the multifaceted dimensions of meaning currently lack a formal characterization. To address our primary question, it is necessary to develop an empirical method to characterize the space of concepts.

We arrive at such a measure by noting that translations uncover the alternate ways that languages partition meanings into words. Many words are polysemous (i.e., they have more than one meaning); thus, they refer to multiple concepts to the extent that these meanings or senses can be individuated (15). Translations uncover instances of polysemy where two or more concepts are fundamentally different enough to receive distinct words in some languages, yet similar enough to share a common word in other languages. The frequency with which two concepts share a single polysemous word in a sample of unrelated languages provides a measure of semantic similarity between them.

We chose an unbiased sample of 81 languages in a phylogenetically and geographically stratified way, according to the methods of typology and universals research (12, 16–18) (*SI Appendix*, section I). Our large and diverse sample of languages allows us to avoid the pitfalls of research based solely on WEIRD societies. Using it, we can distinguish the empirical patterns we detect in the linguistic data as contributions arising from universal conceptual structure from those contributions arising from artifacts of the speakers' history or way of life.

Significance

Semantics, or meaning expressed through language, provides indirect access to an underlying level of conceptual structure. To what degree this conceptual structure is universal or is due to properties of cultural histories, or to the environment inhabited by a speech community, is still controversial. Meaning is notoriously difficult to measure, let alone parameterize, for quantitative comparative studies. Using cross-linguistic dictionaries across languages carefully selected as an unbiased sample reflecting the diversity of human languages, we provide an empirical measure of semantic relatedness between concepts. Our analysis uncovers a universal structure underlying the sampled vocabulary across language groups independent of their phylogenetic relations, their speakers' culture, and geographic environment.

Author contributions: H.Y., E.S., C.M., J.F.W., W.C., and T.B. designed research; H.Y., L.S., E.S., C.M., J.F.W., I.M., and T.B. performed research; L.S. and W.C. collected the data; H.Y., E.S., C.M., J.F.W., I.M., W.C., and T.B. analyzed data; and H.Y., E.S., C.M., W.C., and T.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. Email: visang@santafe.edu or tanmoy@lanl.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1520752113/-/DCSupplemental.

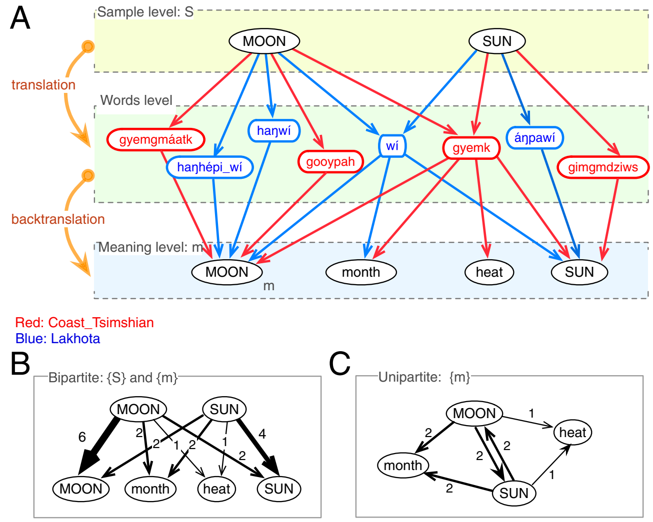


Fig. 1. Schematic figure of the construction of semantic networks. (*A*) Bipartite semantic network constructed through translation (links from the first layer to the second layer) and back-translation (links from the second layer to the third layer) for the cases of MOON and SUN in two American languages: Coast Tsimshian (red links) and Lakhota (blue links). We write the starting concepts from the Swadesh list (SUN, MOON) in capital letters, whereas other concepts that arise through translation (month, heat) are in written in lowercase letters. (*B*) We link each pair of concepts with a weight equal to the number of translation–back-translation paths. (*C*) Resulting weighted graph. More methodological information can be found in *SI Appendix*, section II.

There have been several cross-linguistic surveys of lexical polysemy, and its potential for understanding historical changes in meaning (19) in domains such as body parts (20), cardinal directions (21), perception verbs (22), concepts associated with fire (23), and color metaphors (24). We add a new dimension to this existing body of research by using polysemy data from a systematically stratified global sample of languages to measure degrees of semantic similarity between concepts.

Our cross-linguistic study starts with a subset of concepts from the Swadesh list (25–28). Most languages express these concepts using single words. From the list, we chose 22 concepts that refer to material entities (e.g., STONE, EARTH, SAND, ASHES), celestial objects (e.g., SUN, MOON, STAR), natural settings (e.g., DAY, NIGHT), and geographic features (e.g., LAKE. MOUNTAIN) rather than body parts, social relations, or abstract concepts. The chosen concepts are not defined a priori with respect to culture, perception, or the self; yet, familiarity and experience with them are influenced by the physical environment that speakers inhabit. Therefore, any claim of universality of lexical semantics needs to be demonstrated in these domains first. The detailed criteria of data selection are elaborated in *Materials and Methods* and *SI Appendix*, section I.

Constructing Semantic Network from Translations

We represent semantic relations obtained from dictionary translations of the chosen concepts as a network. Two meanings are linked if they can be reached from one another by a translation into some language and then back. The link is weighted by the number of such paths (of length 2), that is, the number of polysemous words that represent both meanings (details are provided in *Materials and Methods*). Fig. 1 illustrates the construction with examples from two languages. Translating the word SUN into Lakhota results in *wí* and *áŋpawí*. Although the latter picks up no other meaning, *wí* is polysemous: it possesses additional meanings of MOON and month, so they are linked to SUN in the network. A similar polysemy is observed in Coast Tsimshian, where *gyemk*, the translation of SUN, also means heat, thus providing a link between SUN and heat. We write the initial Swadesh concepts (SUN and MOON in this example) in capital letters, whereas other concepts that arise through translations (month and heat here) are written in lowercase letters. We restrict our study to the neighborhood of the initial Swadesh concepts, so further translations of these latter concepts are not followed.

With this approach, we can construct a semantic network for each individual language. It is conceivable, however, that a group of languages bears structural resemblances as a result of the speakers' sharing common historical or environmental features. A link between SUN and MOON, for example, recurs in both languages illustrated in Fig. 1, yet does not appear in many other languages, where other links are seen instead. Thus, for example, SUN is linked to divinity and time in Japanese and to thirst and DAY/DAYTIME in the Khoisan language !Xóõ. The question is then the degree to which these semantic networks are similar across language groups, reflecting universal conceptual structure, and the extent to which they are sensitive to cultural or environmental variables, such as phylogenetic history, climate, geography, or the presence of a literary tradition. We test such questions by constructing aggregate networks from groups of languages that share a common cultural and environmental property and comparing these networks between different language groups.

Semantic Clusterings

As a point of comparison for the networks obtained from such groups of languages, we show the network obtained from the entire set of languages in Fig. 2 and *SI Appendix*, Fig. S6, displaying only the links that appear more than once. This network exhibits the broad topological structure of polysemies observed in our data. It reveals three almost disconnected clusters of concepts that are far more prone to polysemy within each cluster than between them. These clusters admit a natural semantic interpretation. Thus, for example, the semantically most uniform cluster, colored in blue, includes concepts related to water. A second, smaller cluster, colored in yellow, groups concepts related to solid natural materials (e.g., STONE/ROCK, MOUNTAIN) and associated landscape features (e.g., forest, clearing, highlands). The third cluster, colored in red, is more diverse, containing terrestrial terms (e.g., field, floor, ground, EARTH/SOIL), celestial objects [e.g., CLOUD(S), SKY, SUN, MOON], and units of time (e.g., DAY, NIGHT, YEAR). Although the clustering is strong, there do exist rare polysemies that occur only once in our dataset (and are thus not displayed in Fig. 2) connecting the three clusters. Thus, for example, CLOUD(S) is polysemous with lightning in Albanian, whereas the latter is polysemous with STONE/ROCK in !Xóõ, and whereas holy place is a polysemy for MOUNTAIN in Kisi, it is instead polysemous with LAKE in Wintu. The individual networks including such weak links can be accessed in our web-based platform (29).

The links defining each of the three clusters can be understood in terms of well-known kinds of polysemies: metonymies (polysemy between part and whole) and commonly found semantic extension to hypernyms (more general concepts), hyponyms (more specific concepts), and cohyponyms (specific concepts belonging to the same category). The first cluster contains both liquid substances and topographic features metonymically related to water. The substance polysemies in this cluster are various liquids, cohyponyms of WATER. The topographic polysemies (e.g., LAKE, RIVER) are also linked as cohyponyms under "body of water" and "flowing water." Similarly, in the third cluster, the bridge between the terrestrial and celestial components is provided by the hyponyms of "granular aggregates," which span both the terrestrial EARTH/SOIL, DUST, and SAND and the airborne SMOKE and CLOUD(S).

Evidence for Universal Semantic Structure

The semantic network across languages reveals a universal set of relationships among these concepts that possibly reflects human

ANTHROPOLOGY

PSYCHOLOGICAL AND COGNITIVE SCIENCES

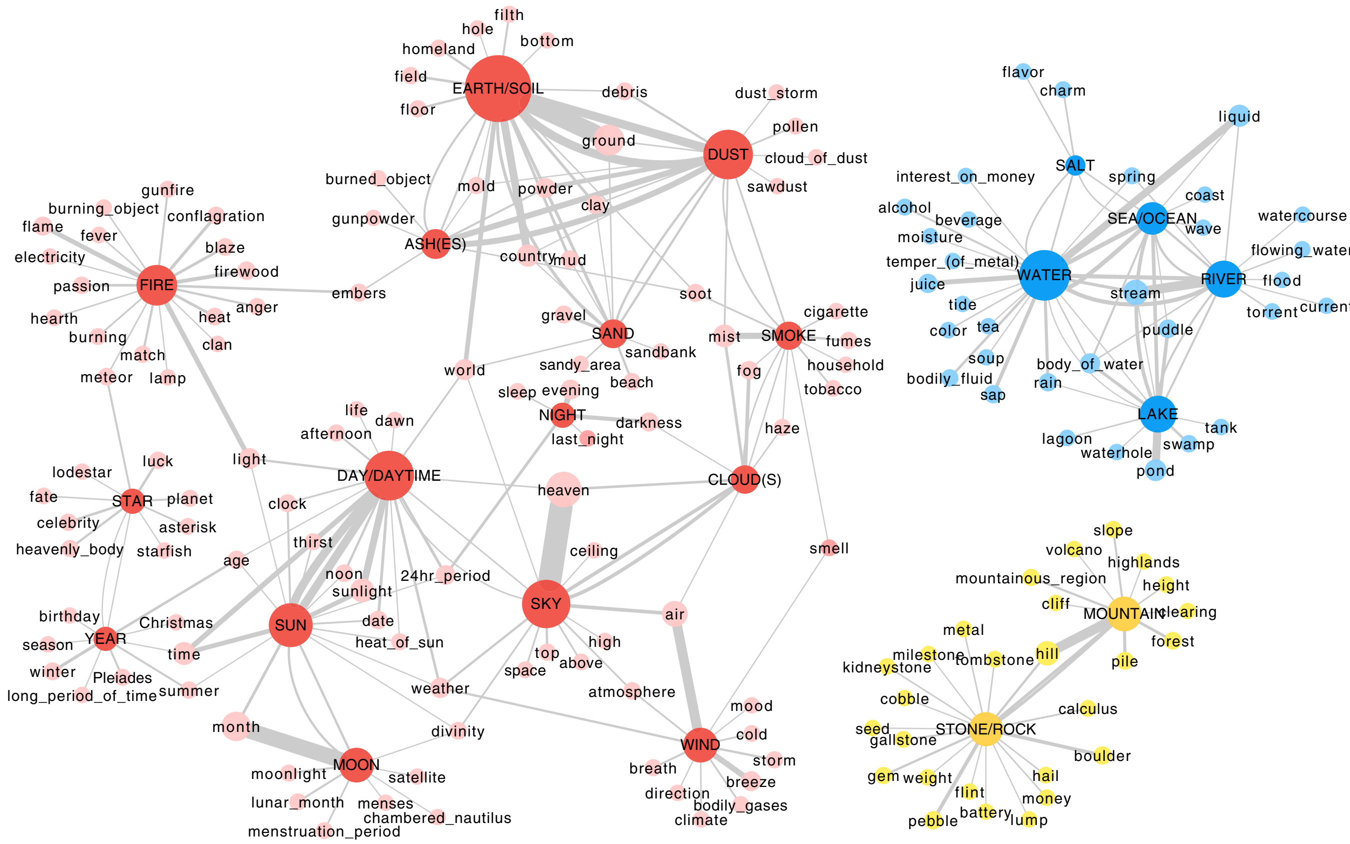


Fig. 2. Semantic network inferred from polysemy data. Concepts are linked when polysemous words cover both concepts. Swadesh words (the starting concepts) are capitalized. The size of a node and the width of a link to another node are proportional to the number of polysemies associated with the concept and with the two connected concepts, respectively. Links whose weights are at least 2 are shown, and their directions are omitted for simplicity. The thick link from SKY to heaven, for example, shows that a large number of words in various languages have both SKY and heaven as meanings. Three distinct clusters, colored in red, blue, and yellow, are identified. These clusters may indicate a set of relationships among concepts that reflects a universal human conceptual structure in these domains.

conceptualization of these semantic domains (8, 12, 30). Alternatively, it has been postulated that such semantic relations are strongly influenced by the physical environment that human societies inhabit (31).

To address this question, we group the languages by various factors (*SI Appendix*, Table SIII) comprising the geography, climate, or topography of the region where they are spoken, and the presence or absence of a literary tradition in them, and we test the effect of these factors on the semantic network. We measured the similarity between these groups' semantic networks in several ways. First, we measured the correlation between the commute distances (32) between nearby concepts (*Materials and Methods* and *SI Appendix*, section III A 1). Then, to compare these networks' large-scale structure, we clustered the concepts in each network hierarchically as a dendrogram (*SI Appendix*, Fig. S8) and compared them using two standard tree metrics (33–35): the triplet distance (D_{triplet}) and the Robinson–Foulds distance (D_{RF}) (*Materials and Methods* and *SI Appendix*, section III A). To test whether these networks are more similar than what we would expect by chance, we performed bootstrap experiments, where we compared each network with the one where the concepts were randomly permuted (*SI Appendix*, section III). As shown by the p_1 values in Fig. 3, in every case, the networks of real language groups are far more similar to each other than to these randomly permuted networks, allowing us to reject decisively the null hypothesis that these semantic networks are completely uncorrelated (statistical details are provided in *SI Appendix*, section III B).

All these tests thus establish that different language groups do indeed have semantic structure in common. To explore this universal semantic structure further, we tested a null hypothesis at the other extreme, that cultural and environmental variables have no effect on the semantic network. For this purpose, we performed a different kind of bootstrap experiment, where we replaced each language group with a random sample of the same size from the set of languages. As denoted by p_2 in Fig. 3, we find that, with rare exceptions, there is no statistical support (*SI Appendix*, section III B) for the hypothesis that the differences between the language groups studied are any larger than between random groups of the same size. This fact means that the impacts of cultural and environmental factors are weaker than what can be established with our dataset; thus, our results are consistent with the hypothesis that semantic clustering structure is independent of culture and environment in these semantic domains.

Heterogeneity of the Semantic Network

The universal semantic network shown in Fig. 2 is heterogeneous in both node degrees and link weights. The numbers of polysemies involving individual meanings are uneven, possibly trending toward a heavy-tailed distribution (Fig. 4). This distribution indicates that concepts have different tendencies of being polysemous. For example, EARTH/SOIL has more than 100 polysemies, whereas SALT has only a few.

Interestingly, we find that this heterogeneity is also universal: The numbers of polysemies of the various concepts that we studied in any two languages are strongly correlated with each other. This correlation holds despite the observation that the languages differ in the overall magnitude of polysemy, so that the same concepts are far more polysemous in some languages than in others (*SI Appendix*, Fig. S2). In fact, a simple formula predicts the number of polysemies, n_{SL}, involving sense S in language L rather well (*SI Appendix*, Fig. S9):

$$n_{SL}^{\text{model}} \equiv n_S \times \frac{n_L}{N}, \qquad [1]$$

where n_S is the number of polysemies involving sense S in the aggregate network from all languages, N is the total number of polysemies in this aggregate network, and n_L is the number of polysemies in the language L. This formula is exactly what we would expect (*Materials and Methods* and *SI Appendix*, section II B) if each language randomly and independently draws a subset of polysemies for each concept S from the universal aggregate network, which we can identify as an underlying "universal semantic space" (*SI Appendix*, Fig. S5). The data for only three concepts, MOON, SUN, and ASHES, deviate from this linear pattern by more than the expected sampling errors ($p \approx 0.01$) in that they display an initial rapid increase in n_{SL} with n_L, followed by a saturation or slower increase at larger values of n_L (*SI Appendix*, Fig. S10). These deviations can be accommodated using a slightly more complicated model described in *SI Appendix*, section IV.

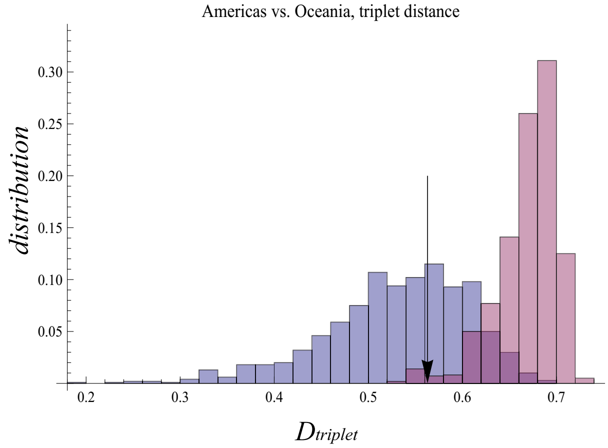
Discussion

We propose a principled method to construct semantic networks linking concepts via polysemous words identified by cross-linguistic dictionaries. Based on the method, we found overwhelming evidence that the semantic networks for different groups share a large amount of structure in common across geographic and cultural differences. Indeed, our results are consistent with the hypothesis that cultural and environmental factors have little statistically significant effect on the semantic network of the subset of basic concepts studied here. To a large extent, the semantic network appears to be a human universal: For instance, SEA/OCEAN and SALT are more closely related to each other than either is to SUN, and this pattern is true for both coastal and inland languages.

These findings have broad implications. Universal structures in lexical semantics such as we observe can greatly aid reconstruction of human history using linguistic data (37, 38). Much progress has been made in reconstructing the phylogenies of word forms from known cognates in various languages, thanks to the ability to measure phonetic similarity and our knowledge of the processes of sound change. The relationship between semantic similarity and semantic shift, however, is still poorly understood. The standard view in historical linguistics is that any meaning can change to any other meaning (39, 40), and no constraint is imposed on what meanings can be compared with detect cognates (41). In contrast to this view, we find that at least some similarities occur in a heterogeneous and clustered fashion.

Previous studies (9, 19–21, 23, 24, 42–45) have investigated the presence or absence of universality in how languages structure the lexicon in a few semantic domains dealing with personal items like body parts, perceptual elements like color metaphors, and cultural items like kinship relations. In this work, we study instead the domain of celestial and landscape objects that one may a priori expect to be strongly affected by the environment. We find, however, that the semantic networks on which these natural objects lie are universal. It is generally accepted among historical linguists that language change is gradual: Over historical time, words gain meanings when their use is extended to similar meanings and lose meanings when another word is extended to the first word's meaning. If such transitional situations are common among polysemies, then the meaning shifts in this domain are likely to be equally universal, and the observed weights on different links of the semantic network reflect the

A



B

Language groups	r	p_1	p_2	D_{triplet}	p_1	p_2	D_{RF}	p_1	p_2
Region									
Americas vs. Eurasia	0.13	0.007	0.003†	0.59	0.06	0.20	30	0*	0.42
Americas vs. Africa	0.73	0*	0.98	0.59	0.02	0.20	30	0*	0.46
Americas vs. Oceania	0.25	0*	0.06	0.56	0.02	0.38	26	0*	0.87
Eurasia vs. Africa	0.41	0*	0.25	0.52	0*	0.62	32	0.01	0.31
Eurasia vs. Oceania	0.27	0*	0.08	0.60	0.06	0.16	28	0*	0.78
Africa vs. Oceania	0.59	0*	0.78	0.46	0*	0.82	34	0.01	0.18
Climate									
Humid vs. Cold	0.59	0*	0.84	0.58	0.04	0.17	28	0*	0.49
Humid vs. Arid	0.07	0.20	0.005†	0.61	0.12	0.10	34	0.01	0.13
Cold vs. Arid	0.33	0*	0.13	0.50	0.01	0.71	30	0*	0.53
Topography									
Inland vs. Coastal	0.74	0*	0.97	0.59	0.04	0.14	28	0*	0.39
Literacy									
Literacy vs. No literacy	0.34	0*	0.08	0.49	0.01	0.62	28	0*	0.45

Fig. 3. (*A*) Illustration of our bootstrap experiments. The D_{triplet} between the dendrograms of the Americas and Oceania is 0.56 (indicated by the downward arrow) (33, 34). This value sits at the very low end of the distribution of distances generated by randomly permuted networks (the red-shaded profile on the right), but it is well within the distribution that we obtain by resampling random groups from the set of languages (the blue-shaded profile on the left). This fact gives strong evidence that each pair of groups shares an underlying semantic network, and that the differences between them are no larger than would result from random sampling (details are provided in *Materials and Methods*). Therefore, these two language groups are much more closely related than if concepts were permuted randomly, showing they share a common semantic structure, but they are roughly as related as any pair of language groups of these sizes, suggesting that the geographic and cultural difference between them have little effect on their structure. (*B*) Comparing distance metrics, the Pearson correlation (r) between commute distances (32) on the semantic networks of groups and the D_{triplet} and D_{RF} among the corresponding dendrograms (33–35), on two bootstrap experiments to obtain p_1 (Mantel test or randomly permuted dendrograms) and p_2 (surrogate groups). The p_1 values for the former bootstrap (p_2 values for the latter) are the fraction of 1,000 bootstrap samples whose distances are smaller (larger) than the observed distance. In either case, 0^* denotes a value below 0.001 (i.e., no bootstrap sample satisfied the condition). The Mantel test used 999 replicates for Pearson correlations to calculate p_1 values, and 99 bootstrap samples (or 999 when marked with †) were used for p_2. Given that we make 11 independent comparisons for any quantity, a Bonferroni-corrected (36) significance threshold of $p_{1,2} = 0.005$ is appropriate for a nominal test size of $p = 0.05$ (more extensively discussed in *SI Appendix*, section III B).

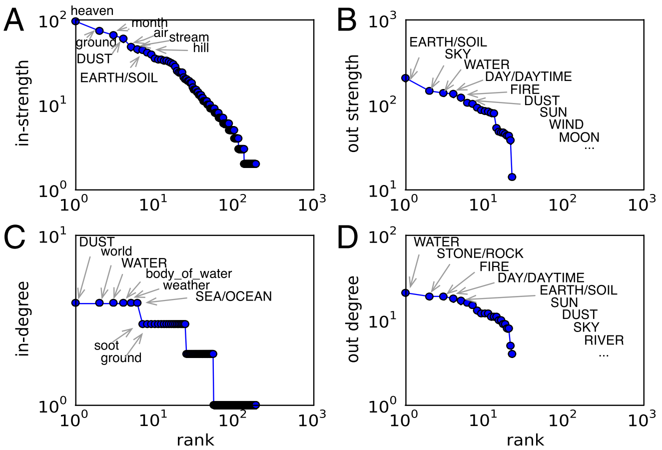


Fig. 4. Rank plot of concepts in descending order of their strengths (total weight of links) and degrees (number of links) shown in Fig. 2. Entries from the initial Swadesh list are distinguished with capital letters. (*A*) In-strengths of concepts: sum of weighted links to a node. (*B*) Out-strengths of Swadesh entries: sum of weighted links from a Swadesh entry. (*C*) In-degree of the concepts: number of unweighted links to a node. (*D*) Out-degree of Swadesh entries: number of unweighted links from a node. A node strength in this context indicates the total number of polysemies associated with the concept in 81 languages, whereas a node degree means the number of distinct concepts associated with the node regardless of the number of synonymous polysemies associated with it. The word "heaven," for example, has the largest number of polysemies, but most of them are with SUN, so that its degree is only three.

probabilities of words to be in transition. As such, semantic shifts can be modeled as diffusion in the conceptual space, or along a universal semantic network, and thus our constructed networks can serve as important input to methods of inferring cognates.

Our results are obtained from detailed typological studies from a sample of the world's languages. We chose to collect data manually from printed dictionaries. This approach ensures that our sample is unbiased and representative of the variation known among languages but foregoes the large sample size that online digital resources offer, because these data are dominated by a few languages from developed countries with long-established writing systems and large speaker populations. We find, however, that the patterns of polysemy in our data have little correlation with environmental, social, and other linguistic attributes of the language. This consistency across language groups suggests that languages for which digital resources are available are likely to produce networks similar to those networks created from the full sample. Therefore, the semantic network constructed here can be extended with more extensive data from online dictionaries and digital corpora by automated means (46). In such analysis with digitally available resources, one can examine if patterns of polysemy could be shared among more closely related language groups than the genus level, and if universality holds for other semantic domains.

Materials and Methods

Polysemy Data. High-quality bilingual dictionaries between a sample of languages and well-known European languages were used to identify polysemies. The samples of 81 languages were selected to be phylogenetically and geographically diverse, covering many low-level language families or genera (16–18, 31) (*SI Appendix*, section I B). The concepts studied were taken from the Swadesh list (25), because these concepts are likely to have historically stable single-word representation in many languages. The domain of study was chosen to extend the existing body of cross-linguistic surveys of lexical polysemy (19–24), and its potential for understanding historical changes in meaning (19) (*SI Appendix*, section I).

We use multiple modern European languages (English, Spanish, French, German, or Russian) interchangeably as semantic metalanguages because sufficiently high-quality bilingual dictionaries were not available in any one of these languages (*SI Appendix*, sections I C and I D). Polysemies were then identified by looking up the translations (and back-translations) of each of the 22 concepts to be studied (*SI Appendix*, section I A) in each language in our sample. All translations (i.e., all synonyms) were retained. The semantic metalanguages themselves sometimes display polysemies: English "day," for example, expresses both DAYTIME and 24HR PERIOD. In our chosen domain, however, the metalanguage polysemy did not create a problem because the lexicographer usually annotates the translation sufficiently. *SI Appendix*, section I elaborate the bases for the procedure and methodology.

Mantel Test. The commute distance between two nodes in a network is the expected number of steps it takes a random walker to travel from one node to another and back, when the probability of a step along a link is proportional to its link weight (32). We connect all word senses with a small weight to provide a finite, although large, distance between disconnected components. To avoid the effect of this modification, the final calculation excludes distances larger than the number of nodes. We then compare the Pearson correlation, r^*, of the commute distances between networks of empirical language groups with the distribution of r that is derived from the bootstrap experiments. The first bootstrap performs the Mantel test (47) by randomly permuting nodes (word senses) of the observed network (p_1), and the second bootstrap compares random groups of languages of the same size (p_2). These tests are carried out for classification by each of the following four variables: geography (Americas, Eurasia, Africa, or Oceania), climate (humid, cold, or arid), topography (inland or coastal), and literary tradition (presence or absence). The language groups and their sizes are listed in *SI Appendix*, Table SIII, and the details of bootstrap methods are provided in *SI Appendix*, section III A 1. All these calculations were done using the statistical package R (48–51).

Hierarchical Clustering Test. A hierarchical spectral algorithm clusters the Swadesh word senses. Each sense i is assigned to a position in $\mathbb{R}^n$ based on the ith components of the n eigenvectors of the weighted adjacency matrix. Each eigenvector is weighted by the square of its eigenvalue and clustered by a greedy agglomerative algorithm to merge the pair of clusters having the smallest Euclidean distance between their centers of mass, through which a binary tree or dendrogram is constructed (*SI Appendix*, Fig. S8).

The distance between the dendrograms obtained from each pair of language groups is measured by two standard tree metrics. The D_{triplet} (33, 34) is the fraction of the $\binom{n}{3}$ distinct triplets of senses that are assigned a different topology in the two trees (i.e., those triplets for which the trees disagree as to which pair of senses are more closely related to each other than they are to the third). The D_{RF} (35), is the number of "cuts" on which the two trees disagree, where a cut is a separation of the leaves into two sets resulting from removing an edge of the tree.

For each pair of groups, we perform two types of bootstrap experiments. First, we compare the distance between their dendrograms with the distribution of distances we would see under a hypothesis that the two groups have no shared lexical structure. Were this null hypothesis true, the distribution of distances would be unchanged under the random permutation of the concepts at the leaves of each tree, despite holding fixed the topologies of the dendrograms. Comparing the observed distance against the resulting distribution gives a P value, called p_1 in Fig. 3. These P values are small enough to reject decisively the null hypothesis. Indeed, for most pairs of groups the D_{RF} is smaller than the distance observed in any of the 1,000 bootstrap trials ($P \lesssim 0.001$), marked as 0^* in the table. These small P values give overwhelming evidence that the hierarchical clusters in the semantic networks have universal aspects that apply across language groups.

In the second bootstrap experiment, the null hypothesis is that the nonlinguistic variables, such as culture, climate, and geography, have no effect on the semantic network, and that the differences between language groups simply result from random sampling: For instance, the similarity between the Americas and Eurasia is what one would expect from any disjoint groups of the 81 languages of given sizes 29 and 20, respectively. To test this null hypothesis, we generate random pairs of disjoint language groups with the same sizes as the groups in question and measure the distribution of their distances. The P values, called p_2 in Fig. 3, are not small enough to reject this null hypothesis. Thus, at least given the current dataset, there is no statistical distinction between random sampling and empirical data, further supporting our claims of universality of conceptual structure (*SI Appendix*, section III A).

Null Model of Degree Distributions. The simplest model of degree distributions assumes no interaction between concept and languages. The number of polysemies of concept S in language L, that is, n_{SL}^{model}, is linearly proportional to both the tendency of the concept to be polysemous and the tendency of the language to distinguish word senses. These tendencies are estimated from the marginal distribution of the observed data as the fraction of polysemy associated with the concept, $p_S^{\mathrm{data}} = n_S^{\mathrm{data}}/N$, and the fraction of polysemy in the language, $p_L^{\mathrm{data}} = n_L^{\mathrm{data}}/N$, respectively. The model, then, can be expressed as $p_{SL}^{\mathrm{model}} = p_S^{\mathrm{data}} p_L^{\mathrm{data}}$, a product of the two.

The Kullback–Leibler (KL) divergence is a standard measure of the difference between an empirical distribution, such as $p_{SL}^{\mathrm{data}} \equiv n_{SL}^{\mathrm{data}}/N$, and a theoretical prediction, p_{SL}^{model} (52, 53). This distance is expressed as:

$$D\left(p_{SL}^{\mathrm{data}} \| p_{SL}^{\mathrm{model}}\right) \equiv \sum_{S,L} p_{SL}^{\mathrm{data}} \log \left(p_{SL}^{\mathrm{data}} \middle/ p_{SL}^{\mathrm{model}}\right).$$

We evaluated the statistical significance of the differences between our model predictions and the experimental degree distribution by comparing the observed KL divergence with the one expected under multinomial sampling with probability p_{SL}^{model}. The P value was calculated as the area under the expected distribution to the right of the observed value (details are provided in *SI Appendix*, section IV).

ACKNOWLEDGMENTS. We thank Ilia Peiros, George Starostin, Petter Holme, and Laura Fortunato for helpful comments. H.Y. acknowledges support from Complex Agent Based Dynamic Networks (CABDyN) Complexity Center and the support of research grants from the National Science Foundation (Grant SMA-1312294). W.C. and L.S. acknowledge support from the University of New Mexico Resource Allocation Committee. T.B., J.F.W., E.S., C.M., and H.Y. acknowledge the Santa Fe Institute and the Evolution of Human Languages program. The project and C.M. are also supported, in part, by the John Templeton Foundation.

1. Whorf BL (1956) *Language, Thought and Reality: Selected Writing* (MIT Press, Cambridge, MA).
2. Fodor JA (1975) *The Language of Thought* (Harvard Univ Press, New York).
3. Wierzbicka A (1996) *Semantics: Primes and Universals* (Oxford Univ Press, Oxford, UK).
4. Lucy JA (1992) *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis* (Cambridge Univ Press, Cambridge, UK).
5. Levinson SC (2003) *Space in Language and Cognition: Explorations in Cognitive Diversity* (Cambridge Univ Press, Cambridge, UK).
6. Choi S, Bowerman M (1991) Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition* 41(1-3):83–121.
7. Majid A, Boster JS, Bowerman M (2008) The cross-linguistic categorization of everyday events: a study of cutting and breaking. *Cognition* 109(2):235–250.
8. Croft W (2010) Relativity, linguistic variation and language universals. *CogniTextes* 4:303–307.
9. Witkowski SR, Brown CH (1978) Lexical universals. *Annu Rev Anthropol* 7:427–451.
10. Evans N, Levinson SC (2009) The myth of language universals: Language diversity and its importance for cognitive science. *Behav Brain Sci* 32(5):429–448, discussion 448–494.
11. Comrie B (1989) *Language Universals and Linguistic Typology* (Univ of Chicago Press, Chicago), 2nd Ed.
12. Croft W (2003) *Typology and Universals* (Cambridge Univ Press, Cambridge, UK), 2nd Ed.
13. Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? *Behav Brain Sci* 33(2-3):61–83, discussion 83–135.
14. Shopen T, ed (2007) *Language Typology and Syntactic Description* (Cambridge Univ Press, Cambridge, UK), 2nd Ed.
15. Croft W, Cruse DA (2004) *Cognitive Linguistics* (Cambridge Univ Press, Cambridge, UK).
16. Bell A (1978) Language samples. *Universals of Human Language, Method and Theory*, eds Greenberg JH, Ferguson CA, Moravcsik EA (Stanford Univ Press, Palo Alto, CA), Vol 1, pp 123–156.
17. Rijkhoff J, Bakker D (1998) Language sampling. *Linguistic Typology* 2(3):263–314.
18. Rijkhoff J, Bakker D, Hengeveld K, Kahrel P (1993) A method of language sampling. *Stud Lang* 17(1):169–203.
19. Koptjevskaja-Tamm M, Vanhove M, eds (2012) New directions in lexical typology. *Linguistics* 50(3):373–743.
20. Brown CH (1976) General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *Am Ethnol* 3(3):400–424.
21. Brown CH (1983) Where do cardinal direction terms come from? *Anthropological Linguistics* 25(2):121–161.
22. Viberg Å (1983) The verbs of perception: A typological study. *Linguistics* 21(1): 123–162.
23. Evans N (1992) Multiple semiotic systems, hyperpolysemy, and the reconstruction of semantic change in Australian languages. *Diachrony Within Synchrony: Language History and Cognition*, eds Kellermann G, Morrissey MD (Peter Lang, Frankfurt, Germany).
24. Derrig S (1978) Metaphor in the color lexicon. *Chicago Linguistic Society, The Parasession on the Lexicon*, eds Farkas D, Jacobsen WM, Todrys KW (The Society, Chicago), pp 85–96.
25. Swadesh M (1952) Lexico-statistical dating of prehistoric ethnic contacts. *Proc Am Philos Soc* 96(4):452–463.
26. Kessler B (2001) *The Significance of Word Lists* (Center for the Study of Language and Information, Stanford, CA).
27. Lohr M (1998) Methods for the genetic classification of languages. PhD dissertation (University of Cambridge, Cambridge, UK).
28. Lohr M (2000) New approaches to lexicostatistics and glottochronology. *Time Depth in Historical Linguistics*, eds Renfrew C, McMahon AMS, Trask RL (McDonald Institute for Archaeological Research, Cambridge, UK), pp 209–222.
29. Youn H (2015) Semantic Network (polysemy) Available at hyoun.me/language/index.html. Accessed December 22, 2015.
30. Vygotsky L (2002) *Thought and Language* (MIT Press, Cambridge, MA).
31. Dryer MS (1989) Large linguistic areas and language sampling. *Stud Lang* 13(2):257–292.
32. Chandra AK, Raghavan P, Ruzzo WL, Smolensy R, Tiwari P (1996) The electrical resistance of a graph captures its commute and cover times. *Comput Complex* 6(4): 312–340.
33. Critchlow DE, Pearl DK, Qian CL (1996) The triples distance for rooted bifurcating phylogenetic trees. *Syst Biol* 45(3):323–334.
34. Dobson AJ (1975) *Comparing the Shapes of Trees, Combinatorial Mathematics III* (Springer, New York).
35. Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Math Biosci* 53(1-2): 131–147.
36. Bonferroni CE (1936) *Teoria statistica delle classi e calcolo delle probabilità. Issue 8 of Pubblicazioni del R Instituto Superiore di Scienze Economiche e Commerciali di Firenze* (Libreria internazionale Seeber, Florence, Italy), pp 3–62.
37. Dunn M, Greenhill SJ, Levinson SC, Gray RD (2011) Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345):79–82.
38. Bouckaert R, et al. (2012) Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097):957–960.
39. Fox A (1995) *Linguistic Reconstruction: An Introduction to Theory and Method* (Oxford Univ Press, Oxford, UK).
40. Hock HH (1986) *Principles of Historical Linguistics* (Mouton de Gruyter, Berlin).
41. Nichols J (1996) The comparative method as heuristic. *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, eds Durie M, Ross M (Oxford Univ Press, New York).
42. Fox R (1967) *Kinship and Marriage: An Anthropological Perspective* (Cambridge Univ Press, Cambridge, UK).
43. Greenberg JH (1949) The logical analysis of kinship. *Philos Sci* 16(1):58–64.
44. Greenberg JH (1966) *Language Universals, with Special Reference to Feature Hierarchies*, Janua Linguarum, Series Minor 59 (Mouton, The Hague).
45. Parkin R (1997) *Kinship* (Blackwell, Oxford).
46. List J-M, Mayer T, Terhalle A, Urban M (2014) CLICS: Database of Cross-Linguistic Colexifications (Forschungszentrum Deutscher Sprachatlas, Marburg, Germany), Version 1.0. Available at CLICS.lingpy.org. Accessed August 27, 2015.
47. Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27(2):209–220.
48. R Core Team (2015) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna), Available at https://www.R-project.org/. Accessed November 30, 2015.
49. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* (Springer, New York), 4th Ed. Available at www.stats.ox.ac.uk/pub/MASS4. Accessed November 30, 2015.
50. Csardi G, Nepusz T (2006) The igraph software package for complex network research, InterJournal (Complex Systems) 1695. Available at igraph.org/. Accessed November 30, 2015.
51. Dray S, Dufour AB (2007) The ade4 package: Implementing the duality diagram for ecologists. *J Stat Softw* 22(4):1–20.
52. Kullback S, Leibler RA (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22(1):79–86.
53. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).

ANTHROPOLOGY

PSYCHOLOGICAL AND COGNITIVE SCIENCES